Assignment IV – Amos II Database Project

Molecular Biology

1. Goals

The goal of this exercise is to give a practical experience in developing of a small database application using an object-relational database management system. Several of the large commercial relational database vendors have started to introduce such object-oriented extensions in their SQL implementations. In this assignment the students will work with the AMOS II object-relational DBMS, which is a research prototype system being developed at Uppsala Database Laboratory, Uppsala University.

The exercise consists of the development of a small *molecular geometry database application* for storing and retrieving information about molecules by implementing schema definitions and queries in AMOSQL, the query language of AMOS II.

2. Preparations

Download AMOS II from http://user.it.uu.se/~udbl/amos/. Install the system on a PC with Windows.

Write your solutions on paper before testing them out on the AMOS II system.

3. Background reading

o Elmasri/Navathe: chapter 20, 21 and 22

o Padron-McCarthy/Risch: Chapter 16

o All material from the lectures on object-oriented and object-relational databases systems and query languages. See also the lab course web page for additional information.

Information about molecule geometry can for example be found at:
http://www.shef.ac.uk/chemistry/vsepr/

http://www.worldofmolecules.com/.

4. Instructions for the assignment

The exercise consists of 2 parts:

1.) Work through the AMOS II tutorial that is part of the AMOS II archive to download.

2.) Develop a molecular geometry database application to handle information regarding the structure and geometry of classes of molecules. Your test data should include at least the different molecule classes Linear, Trigonal Planar, Tetrahedral, Trigonal Bipyramidal and Octahedral. Additional molecule classes can be selected individually.
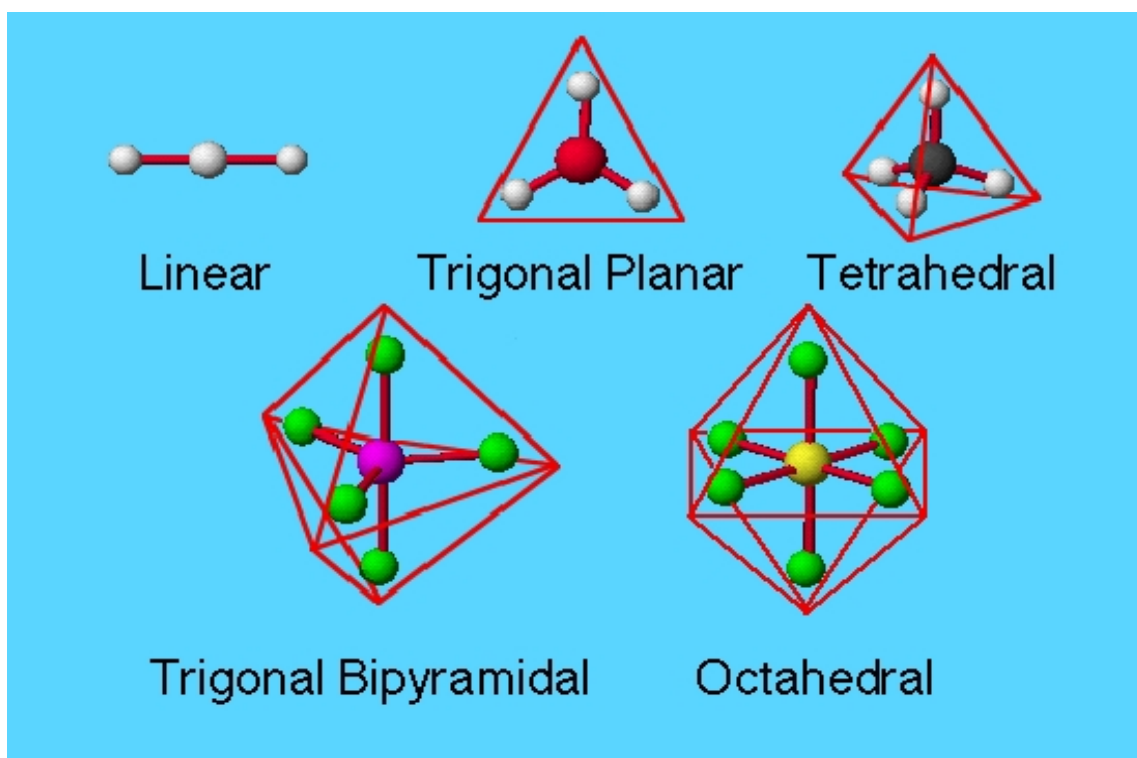
Figure 1. Molecular geometry types.

Number of electron pairs				
Total	Bonding	Nonbonding	Molecule (or ion) shape	Examples
2	2	0	linear	$HgCl_2$, $CuCl_2^-$
3	3	0	triangular planar	BF_3, $HgCl_3^-$
3	2	1	angular	$SnCl_2$, NO_2^-
4	4	0	tetrahedral	CH_4, BF_4^-
4	3	1	trigonal pyramidal angular	NH_3, PF_3
4	2	2	angular	H_2O, ICl_2^+
5	5	0	trigonal bipyramidal	PCl_5, $SnCl_5^-$
5	4	1	irregular tetrahedral	$TeCl_4$, IF_4^+
5	3	2	T-shaped	ClF_3, BrF_3
5	2	3	linear	XeF_2, ICl_2^-
6	6	0	octahedral	SF_6, PF_6^-
6	5	1	square pyramidal	IF_5, SbF_5^{2-}
6	4	2	square planar	BrF_4^-, XeF_4

Table 1. Number of electron pairs in the valence shell of the central atom and the shape of the atom or ion (from Chemistry, 6th ed, by Charles E. Mortimer, Wiley 1986).

The data collected for each molecule class should include information such as name, weight, central atom, number of valence electrons, attached groups, atoms or ions, atom (or ion) radii, number of bonding and nonbonding electrons, bonds, bond types and bond angles. Complement this information, if necessary to complete the assignment. For calculating spatial extensions of molecules, you can use an idealized model treating atoms as rigid spheres with the radius corresponding to their atomic radius.

Typical queries you should be able to answer are:

1. The name and type of molecules.

2. The weight of atoms as well as of molecules.

3. The angle between arbitrary atoms in the molecule.

4. The number of various bond types.

5. Which atoms are part of a certain bond.

6. The bounding volume (spherical) of a molecule.

7. Comparison of data for an unknown molecule with data stored in the database.

8. Matching a molecule of an unknown type with molecule types in the database.

9. Which molecule type does match a bounding sphere with a radius of 2,57 Å(ngström) and a molecule weight of 15u.

10. Which molecule type does match a bounding sphere with a radius of 2,54 Å and a molecule weight of about 18u.

Any additional and interesting queries are welcome. At least two such additional queries should be defined.

5. Handing in

Hand in an overview of the design, i.e. in an EER diagram, including explanations to concepts and symbols, and source code for creating the database. Include solutions with system replies to all the questions in the exercise as a printout of the interaction with AMOS II. This can be done by copying the results from the window where you are running AMOS II to a text file that you print out and hand-in to your assistant. You do not have to hand in the answers of the tutorial exercises.